DOCUMENT RESUME

ED 406 415 TM 026 323

AUTHOR Kennedy, Robert L.
TITLE External Validation of an Assessment Class.
PUB DATE 7 Nov 96
NOTE 9p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Tuscaloosa, AL, November 6-8, 1996).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Content Validity; *Course Content; *Educational Assessment; *Education Majors; Elementary Education; Elementary School Teachers; Evaluation Methods; Higher Education; *Knowledge Base for Teaching; Performance Based Assessment; Student Evaluation; Test Construction; User Needs (Information)

ABSTRACT

A survey was conducted to establish the content validity of the material presented during a diagnostic and evaluative procedures course for elementary education majors. Although the content was consistent with information typical of the course, validation through the opinions of practicing teachers could strengthen its validity with the students. An additional benefit is that class members would have testimonials to the value of the class from sources other than their instructor. The 21-question survey was based on the course content outline. Most questions were of the "Do you use..." or "Do you need to know..." variety, but there were several open-ended questions, including one asking for recommendations. Members of 4 assessment classes over 2 years were required to interview from 3 to 10 teachers, depending on the semester, for a total of 333 practitioners from a wide demographic range. Results indicate that the customary topics, including behavioral objectives, Bloom's taxonomy, and short-answer test items, are being used. Portfolios and performance testing appear to be well-established, while norm-referenced standardized tests appear to be declining in popularity. Most of the teachers had few content ideas to suggest, possibly indicating that the course content is at least adequate for teachers' needs. The teacher survey is attached. (Contains five references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

External Validation of an Assessment Class

Robert L. Kennedy

University of Arkansas, Little Rock

ED 406 415

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

ROBERT L. KENNEDY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

BEST COPY AVAILABLE

Mid-South Educational Research Association

Annual Meeting

November 7, 1996

Bryant Conference Center

Tuscaloosa, Alabama

TM026323

Abstract

The purpose of the survey was to content validate the material presented during a diagnostic and evaluative procedures course for elementary education majors. Although the content was consistent with information typical of the course, validation by teaching practitioners could strengthen its credibility with the students. A serendipitous benefit is that class members would hear testimonials from sources other than their instructor as to the value of the class.

The twenty-question survey form was based on a content outline of the course. Most of the questions were of the "Do you use . . . ?" or "Do you need to know . . . ?" variety, but there were several open-ended questions, including one asking for recommendations. Members of four assessment classes over the past two years were required to interview from three to ten teachers, depending on the semester, for a total of 333 practitioners, coming from a wide demographic range.

The results of the survey indicated that customary topics including behavioral objectives, Bloom's Taxonomy, and short-answer test items are being used. Portfolios and performance testing appear to be well-established while norm-referenced standardized tests seem to be declining in popularity.

External Validation of an Assessment Class

Robert L. Kennedy

Practicing teachers can bring knowledge and experience, not only to the classroom, but also to teacher education and professional development (Dilworth and Imig, 1995; Kjelgaard and Norris, 1994). By reviewing the content in courses specific to teachers, practitioners can share their "practical wisdom" (Kjelgaard and Norris, 1994, p.12) with the participants in the teacher education programs (Dilworth and Imig, 1995). It seems reasonable to ". . . acknowledge the value and power inherent in sharing their various perspectives and openly critiquing them" (Condon and Clyde, 1993, p.73).

One process for involving practitioners in content validating an assessment course is illustrated in this paper. Senior elementary education students in a Diagnostic and Evaluative Procedures class were asked to survey teachers known to them, about the content included in the course. The survey was originally designed in the Fall of 1994 by the instructor and included a list of the content normally included in the class, derived from the text in use at that time (Tuckman, 1988). Although the instructor later switched to Gronlund (1993), the content included in the instrument is still viable since it is typical fare for an introductory assessment course. Students interviewed or distributed surveys (see Appendix) to three to ten teachers, depending on the semester, requesting that they indicate whether information about the content item listed was needed by them to carry out their teaching responsibilities. In some cases, questions were open-ended, requiring more than a "yes" or "no" response, and providing more insight into their "practical wisdom".

Findings

The 333 teachers involved in the study represented most areas of the state, including 13 from other states in the South and Midwest, although the majority (96%) came from central Arkansas. Almost all of the schools were public (95.5%), although 15 were private. The grade level representation was fairly evenly divided among pre-kindergarten and kindergarten (13.5%), first grade (21.6%), second (18.6%), third (19.2%), fourth (23.1%), fifth (21.6%), and sixth (13.5%). (Note that the totals presented in this paper will not always equal 100% because some questions allow for multiple responses from the same teacher.)

The first content question asked, "Do you write behavioral objectives as part of your planning for tests? About two-thirds (69.1%) said they did. Another 30.3% said no or not applicable and 0.6% gave no response. The next questions asked whether the respondents used Bloom's Taxonomy, content outlines, and test-item specifications in their planning for tests. The answers for Bloom's were yes (78.1%), no or not applicable (21.3%) and no response (0.6%); for outlines, yes (54.4%), no or not applicable (43.5%), no response (2.1%); and for specifications, yes (50.8%), no (44.7%), no response (4.5%).

The next group of questions inquired about the types of questions used by the respondents in their tests. For unstructured test items, the answers were yes (76.9%), no (or not applicable) (19.8%), and no response (3.3%); for completion, yes (79.9%), no (18.3%), no response (1.8%); for true-false, yes (56.5%), no (41.1%), no response (2.4%); for two-choice classification, yes (32.1%), no (62.5%), no response (5.4%); for multiple choice, yes (80.5%), no (17.1%), no response (2.4%); for matching, yes (78.4%), no (19.2%), no response (2.4%); and for essay-type, yes (56.8%), no (41.1%), no response (2.1%).

The teachers were next asked, "Do you use performance-type tests?". A positive response came from 82.0% with 16.2% no and 1.8% no response. If the response was "yes", then they were asked what kinds of performance tests were used. The most common responses were writing (49.2% of all respondents), portfolios (37.8%), science projects or experiments (28.5%), and dramatic presentations or acting (16.2%). Other answers included unspecified projects (10.5%), unspecified presentations (5.7%), teacher observations (2.7%), and social studies projects (2.4%).

Teachers were then asked "What means, if any, do you use to insure that your tests have content validity?" Most answered that they matched what they tested with what they taught (30.9%) or that they followed district curricular guides, the text, or other published materials (22.8%). 25.5% did not respond, while 13.5% said they did nothing or content validity was not applicable. To the question, "What means, if any, do you use to build reliability into your tests?", 10.2% each responded with the same remarks, that they matched what they tested with what they taught or that they followed district curricular guides, the text, or other published materials. Nearly 7% (6.9%) said they judged reliability through students' performance on the tests. Most provided no response (45.6%), although 11.1% said they did nothing and 6.0% said reliability was not applicable.

Teachers were also asked if they used standardized tests. Most (76.6%) responded that they did. The rest (23.4%) responded no or provided no response. For teacher-made tests, the teachers were asked which types of test items they most frequently used. Their responses were multiple choice (52.0%), completion (fill-in) (48.0%), matching (39.9%), unstructured (open-ended) (32.1%), essay (24.3%), true-false (24.0%), and two-choice (7.8%). Some (10.8%) did not respond.

The next question was, "How are norm-referenced tests, like the Stanford, useful?" The most common response was "diagnostics" (23.4%) followed by "comparisons" (16.5%) and "placement" (13.2%). Most of the other respondents said either "not useful" (12.3%) or gave no response (18.9%). On the other hand, to the question "What are their drawbacks?", respondents most frequently cited "invalid measure" (19.5%), "too much emphasis on one test" (or one day) (13.2%), some students do not test well (7.5%), "biased" (7.2%), and "too stressful" (6.6%). Other responses included "too hard" (5.4%), "too long" (4.5%), "ignores other student characteristics" (4.2%), and "compares students" (3.6%). No response was given by 25.2% of the respondents.

Similar questions were asked about criterion-referenced tests. The most common responses to the usefulness of these tests were "show progress" (12.6%) "shows strengths and weaknesses" (8.4%). Some of the teachers said that they were not useful (4.5%) or they did not use them

(11.1%) and 33.9% gave no response. Drawbacks to criterion-referenced tests included "invalid measure" (5.7%), "too stressful" (4.8%), and it leads to teaching to the test, too much emphasis on one test, and encourages minimum learning (3.6% each). No response was listed by 52.9% of the teachers.

The teachers were then asked, "What types of standard scores do you use or have need to be able to interpret?" For z-scores, the responses were yes, 15.9%; and no, not applicable, or no response, 84.1%. For T-scores, the responses were yes, 15.6%, no, (or not applicable), or no response, 84.4%; for CEEB (College Board) scores, yes, 3.0%, no or no response, 97.0%; for AGCT (Army General Classification Test) scores, yes, 3.0%, no or no response, 97.0%. Asked whether they needed to be able to interpret stanines, percentile ranks, grade-equivalent scores, standard deviations, and Wechsler Scales, the teachers replied: stanines, yes, 57.7%, no (or not applicable) or no response, 42.3%; percentiles, yes, 83.2%, no or no response, 16.8%; grade-equivalents, yes, 79.6%, no or no response, 20.4%; standard deviations, yes, 54.1%, no, 45.9%; and Wechsler Scales, yes, 33.3%, no, 66.7%.

Finally, the teachers were asked, "Do you have any recommendations concerning the teaching of the course in Diagnostic and Evaluative Procedures in terms of any content or other aspects?" Although most of the respondents provided no suggestions (73.3%), there were a few: "know how to interpret standardized test scores" (4.5%), "more on portfolios" (2.7%), "teach what teachers use" and teach a variety of testing tools (2.4% each), "more on performance tests", "how to report test results to parents", and "how to make tests" (2.1% each).

Summary and Conclusions

For all questions there were some positive responses, indicating that most topics are worth including in a course on Diagnostic and Evaluative Procedures. A couple of notable exceptions were the College Board and Army General Classification Test scores. Only three percent of the teachers indicated a need to be able to interpret these scores, suggesting that these topics could be deleted without harming the course. The majority of teachers are using performance tests, most frequently through writing, portfolios, science projects or experiments, and dramatic presentations. Nearly two-fifths (39%) of the teachers listed no means for insuring content validity in their tests, a very important test trait. Although they may simply not have wanted to take the time to write down their answer, it seems likely that at least some teachers may not know how to insure content validity, or possibly do not understand the concept. Most teachers indicated that they used standardized tests, but did not seem to be satisfied with them, their biggest concern being that the tests may be invalid measures. The teachers did seem to be more satisfied with criterion-referenced tests since far fewer listed disadvantages than for standardized tests. Standard scores were among the concepts least utilized by the respondents. Most of the teachers had few other content ideas to suggest, possibly indicating that the content of the course is at least adequate for their needs.

Several topics involved in test preparation, including content outlines, test-item specifications, and content validity are used by only half of the teachers, yet are commonly included in evaluation texts. There may be a need here to provide greater support or explanation of these concepts to increase their usage, if that is still regarded as desirable. Other topics, including standardized and criterion-referenced tests, do not seem to be particularly appreciated

or valued, although most teachers are using them. These appear to be areas where evaluation teachers need to either reassess the importance of the concepts, or stress their roles in assessment, or both. This survey, then, provided some insight into the range of evaluative concepts actually implemented by practicing teachers. Teachers who contribute to the professional development of students in teacher education may find support here for teaching a wide selection of assessment topics. Inviting practitioners to help inform this discussion paints a realistic picture for both teachers and students, and may be a key to further enhancing the growth of young professionals.

References

Condon, M. W. F., & Clyde, J. A. (1993, December). New roles and relationships in teacher preparation. Equity and Excellence in Education, 26, 3, 73-76.

Dilworth, M. E., & Imig, D. G. (1995). Professional Teacher Development and the Reform Agenda. ERIC Digest. Washington, DC: ERIC Clearinghouse on Teaching and Teacher Education, Office of Educational Research and Improvement. (ERIC Document Reproduction Service No. ED 383 694)

Gronlund, N. E. (1993). How to Make Achievement Tests and Assessments (5th ed.). Boston: Allyn and Bacon.

Kjelgaard, P. A., & Norris, C. A. (1994, April). Teacher training in the professional development model. Implications for students at risk. Computing Teacher, 21, 7, 12-14.

Tuckman, B. W. (1988). Testing for Teachers (2nd ed.). San Diego, CA: Harcourt, Brace, Jovanovich.

Student_____

EDEN 4205 Teacher Survey
Diagnostic and Evaluative Procedures in Education

I have been asked by one of my instructors at UALR to survey practicing public elementary school teachers for their opinions, from the practitioner's standpoint, about the content which should be taught in the Diagnostic and Evaluative Procedures in Education course. Would you be willing to participate in this survey, if you have not already? Your name will not be used [and should not be written on this form]. (If the response is no, you will need to find another teacher. If the response is yes, please write down the district employing the teacher and the grade level(s) taught.)

District: _____ Grade Level(s): _____

Would you please respond to the following questions to the best of your knowledge and experience. We are trying to gather information to make the assessment class as realistic and meaningful as possible. If you have any additional comments you wish to make, feel free to add them at any time. List any comments to the right or on the back, indicating to which item the comments belong. Be sure that you can explain what each of these items is (See Tuckman if you are unsure.).

1. Do you write behavioral objectives as part of your planning for tests?

2. Do you use Bloom's Taxonomy as part of your planning for tests?

3. Do you write content outlines as part of your planning for tests?

4. Do you write test-item specifications as part of your planning for tests?

5. Which short-answer test-item types do you use in testing?:

   a. unstructured (can be answered by a word, phrase, or number)

   b. completion (fill in an omitted word or phrase)

   c. true-false (yes-no)

   d. two-choice classification

   e. multiple choice

   f. matching

6. Do you use essay-type test items in testing?

7. Do you use performance-type tests?

   If so, what kinds (eg., writing, dramatic presentations, science projects, portfolios)?

8. What means, if any, do you use to insure that your tests have content validity?

9. What means, if any, do you use to build reliability into your tests?

10. Do you use standardized tests?

11. What types of test items (eg., unstructured, completion, true-false, two-choice, multiple choice, matching, essay) do you most frequently use on your teacher-made tests?

12. a. How are norm-referenced tests, like the Stanford, useful?

    b. What are their drawbacks?

13. a. How are criterion-referenced tests, like the MPT, useful?

    b. What are their drawbacks?

14. What types of standard scores do you use or have need to be able to interpret?

    a. z-scores

    b. T-scores

    c. CEEB scores

    d. AGCT scores

15. Do you need to be able to interpret stanine scores?

16. Do you need to be able to interpret percentile ranks?

17. Do you need to be able to interpret grade-equivalent scores?

18. Do you use or need to be able to interpret standard deviations?

19. Do you need to be able to interpret Wechsler Scales?

20. Do you have any recommendations concerning the teaching of the course in Diagnostic and Evaluative Procedures in terms of any content or other aspects?

Thank you very much for your help. Your comments will contribute to the quality of the course. We appreciate your time and thoughts. (Be sure to be enthusiastic in expressing your appreciation. They did you a favor.)